UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001

February 28, 2001

DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM SERIES B-14*

MEMORANDUM FOR Howard Hogan
Chief, Decennial Statistical Studies Division

From: Donna Kostanich DK
Assistant Division Chief, Sampling and Estimation
Decennial Statistical Studies Division

Prepared by: Richard A. Griffin R.G.
Chief, Estimation Staff

Donald J. Malec D.M
Principal Researcher
Statistical Research Division

Subject: Accuracy and Coverage Evaluation: Assessment of Synthetic Assumption

The attached document was prepared, per your request, to assist the Executive Steering Committee on A.C.E. Policy in assessing the data with and without statistical correction.

This report focuses on the Synthetic Assumptions for the 2000 Census Accuracy and Coverage Evaluation Survey. The analysis deals with errors in synthetic estimates computed for geographic areas smaller than post-strata, specifically states and congressional districts.

DSSD Census 2000 Procedures and Operations Memorandum Series B-14*
February 28, 2001

Accuracy and Coverage Evaluation: Assessment of Synthetic Assumption

Richard A. Griffin and Donald J. Malec
U.S. Census Bureau

Table of Contents

Tables

Accuracy and Coverage Evaluation: Assessment of Synthetic Assumption

prepared by Richard A. Griffin and Donald J. Malec

Executive Summary

The A.C.E. estimation methodology produces estimated coverage correction factors for each of the post-strata. These factors are applied or carried down within the post-strata to the census block level. This process is referred to as synthetic estimation. The key assumption underlying this methodology is that the net census coverage, estimated by the coverage correction factor is relatively uniform within the post-strata. Failures of this assumption lead to synthetic error.

It is important to understand that the design underlying the synthetic estimation methodology is directed at correcting for a systematic under or over count in the census. The synthetic estimates will not result in the correction of random counting errors that occur for any entity (blocks tracts, counties, etc). Therefore, the synthetic estimate will not result in extreme changes in small geographic entities, nor will it correct for extreme errors. It is designed to remove the effects of systematic errors so that when small entities are aggregated, systematic and differential coverage errors are corrected.

The Census Bureau is concerned with synthetic error since it is not included directly in the total error model. Furthermore, synthetic error cannot be estimated directly since this would require more sample observations for the A.C.E than practicable. The analysis of the effects of synthetic error are based on the construction of "artificial populations." These are populations that are created with surrogate variables that are known for the entire population, and are developed to reflect the distribution of net coverage error. An analysis of these populations for the effect of synthetic error is the basis on which this otherwise unknown effect is studied.

We assessed the level of bias in synthetic estimates at the state and congressional district levels. This involved defining the components of error in the synthetic estimate, creating artificial populations to estimate one of these components, and estimating the other component by obtaining post-stratum Dual System Estimate levels of bias including correlation bias from the Total Error Model.

What is the synthetic assumption?

The synthetic assumption states that net census coverage does not vary within post-strata. For example, the synthetic assumption implies that net coverage in St. Louis, Missouri in a given post-stratum is the same as net coverage in the same post-stratum but in Milwaukee, Wisconsin.

What are synthetic estimates?

A synthetic estimate of population is the sum over post-strata for a particular geographic area of interest of the post-stratum census coverage correction factor times the post-stratum census count for that area.

What are the components of error in synthetic estimates?

The bias of a synthetic estimate for a geographic area can be split into two components: (1) Bias due to applying the same coverage correction factor to areas with different net coverage and (2) Bias in the Dual System Estimate (DSE) including correlation bias.

What are the components of bias in the census count?

The census counts also suffer from net coverage differences across areas.

How are these components of bias estimated?

The synthetic bias due to differing net coverage is estimated using artificial populations. The bias due to DSE is estimated by obtaining the post-stratum-level bias in the DSE from the Total Error Model and distributing it to small areas in proportion to their census counts. The census bias is estimated using artificial populations.

What is an artificial population?

We want to compare the synthetic estimates and the census counts for geographic areas to the true counts. However, we do not know the true population for a geographic area such as a congressional district. Surrogate variables correlated with gross undercount and/or gross overcount which are available for small areas are used to create artificial populations. The known population counts for these surrogate variables are used to scale post-stratum-level gross undercount and overcount estimates to produce target or true population counts.

Artificial populations, thus, involve surrogate variables, not the real variable of interest. This is a limitation to consider when examining these results.

What is the relative bias in state synthetic estimates of total persons ?

The average relative bias in the state synthetic estimates is about 0.8 percent for all four artificial populations.

What is the absolute relative bias in state synthetic estimates of population shares?

The average relative bias in the state synthetic estimates of population shares is about 0.7 percent for three of the artificial populations and 0.14 percent for the other artificial population.

How does bias in the census count compare with the bias in synthetic estimates at the congressional district level?

The median ratio of absolute census bias to absolute synthetic bias for levels for congressional districts is between about 1.4 and 1.6, meaning that there is more bias in the census count than in the synthetic estimate.

The median ratio of absolute census bias to absolute synthetic bias for estimated shares for congressional districts is between about 1.4 and 2.1, meaning that there is more bias in the census count than in the synthetic estimate.

The loss function analysis does not include a measure of error due to the synthetic assumption. What is the effect of this bias on the loss function results?

For state level estimates using the weighted squared error loss function, all four artificial populations show a negative correction in the census loss minus the A.C.E. loss to correct this bias. This means the loss function analysis overestimates the true gains from adjustment. Since the bias corrections are negative by only small amounts relative to the loss function analysis results, correcting for this bias would not change the loss function results to favor the census.

For congressional district (CD) share estimates, using the equal CD loss function, the bias correction is positive for one of the four artificial populations, and the results of the loss function analysis thus remain favorable to the A.C.E. The other three artificial populations show a negative correction in the census loss minus the A.C.E. loss to correct this bias. However, the bias corrections are negative by only small amounts relative to the loss function analysis results, so correcting for this bias would not change the loss function results to favor the census.

All loss function results cited in this report use the model which includes correlation bias except for Non-Blacks ages 18-29 and use the Gross DSE to distribute target estimates. Using Gross

Undercount to distribute target estimates keeping the correlation bias assumption fixed would produce results of similar magnitude and sign. We did not run alternative correlation bias assumptions; we think these results are reasonable under these alternatives but we are not completely confident of this.

Introduction

The synthetic assumption states that census net coverage does not vary within post-strata. For example, the synthetic assumption implies that census counts in St. Louis, Missouri in a given post-stratum have the same net coverage as the census counts in the same post-stratum but in Milwaukee, Wisconsin. The synthetic assumption within post-strata will permit the Census Bureau to draw conclusions from the A.C.E. sample about the population as a whole, to individuals living in geographic areas smaller than post-strata. The synthetic assumption is necessary to permit correction for small geographic areas based on a sample. This adjustment is only correcting for systematic biases and not local census errors. The error that is introduced when the synthetic assumption does not hold is called synthetic error.

Assessments of the 1990 PES were concerned with the possibility that synthetic error introduced error in the PES, especially for low levels of geography such as blocks. Synthetic error is of greater concern for small areas than for larger geographic aggregations. It is acknowledged that synthetic error will likely result in the population of some blocks being overestimated and the population of other blocks being underestimated; statistical correction is not expected to produce unqualified improvement in the smallest geographic areas, like blocks.

While the accuracy of the A.C.E's synthetic estimates depends on the degree in which net coverage varies within post-strata, it is important to understand that perfectly equal net coverage cannot exist within all post-strata. The Census Bureau's evaluation of synthetic error should focus on whether the variability of net coverage is so great as to prevent an improvement from using the A.C.E. Additionally, the A.C.E. was designed to reduce the variability of net coverage as compared with the 1990 PES. The A.C.E. design has enhanced post-strata, including variables for mail return rate and type of enumeration areas. In addition, the census has net coverage that varies across areas.

This paper presents alternative methods to document and measure synthetic error in the A.C.E. and the effects, if any, these violations had on the overall accuracy of the A.C.E., both numeric and distributive. The two components of error in synthetic estimates are: (1) Synthetic population bias due to applying the same coverage correction factor to areas with different net census coverage and (2) bias in the Dual System Estimate (DSE) including correlation bias. Synthetic bias is measured at the Congressional district and state levels and is compared to error in the census.

Overview of 1990 evaluations

Evaluations of synthetic estimates, using surrogate variables to create artificial populations of population counts have been documented in Fay and Thompson (1993), Freedman and Wachter (1994) and Kim et al. (1995). In particular, Freedman and Wachter (1994) document a number of analyses using artificial populations. They provide estimates of the within post-strata and between post-strata variability; demonstrating within post-strata variability. A loss function analysis on the surrogate variables is also provided by Freedman and Wachter (1994). Although the loss function analysis (on shares) is favorable to the use of the synthetic estimates (based on a census adjustment), it is noted that the assumptions about the representativeness of the artificial populations are tentative and give variable results. In addition, Freedman and Wachter also show that loss function analysis using the synthetic estimate as the target may overstate the advantage of adjustment. This latter shortcoming is corrected to an extent, using some simplifying assumptions, by Fay and Thompson (1993) who perform a loss function analysis that incorporates both the artificial loss function of the synthetic estimator with a loss function that measures the other sources of bias and error in the DSE. In that analysis, the results are mixed. Kim et al. (1995) analyze state effects using both artificial populations and PES data. They also report mixed results. Hengartner and Speed (1993) analyze PES counts at the block level and find heterogeneity beyond post-strata.

Overview of methodology

This section describes the essence of estimating bias in synthetic estimates. There are two components of error in synthetic estimates - synthetic population bias and bias in the DSE including correlation bias. The Appendix provides the mathematical details of the methodology.

Creation of artificial populations

The basic methodology used to estimate the synthetic population bias component of synthetic error is artificial populations.

We use census variables thought to be related to coverage to produce artificial populations. Call these variables surrogates. We use methodology similar to one method suggested by Freedman and Wachter (1994). Adjust one surrogate variable to gross undercount and another to gross overcount. This is done by distributing the post-stratum level gross undercount (gross overcount) proportional to the gross undercount surrogate variable (gross overcount surrogate variable) for the congressional districts (see Appendix). These are added and subtracted to census counts to form an artificial population. Unlike other approaches, this strategy can provide both net over- and under- coverage between local areas within a poststrata. It is possible that the surrogates that are best for gross undercount are different than those that are best for gross overcount.

The surrogate variables considered are:

- Allocations -households with more than a specified amount of item nonresponse (Items include race, Hispanic origin, relationship, sex, and age)
- Number of Non-Mail Returns
- Number of Substitutions -whole-household imputes and/or partial household substitutions
- Number of duplicates added back (late adds)
- Units at basic street address

Allocations, substitutions, multi-unit, and non-mail back were surrogates used by Freedman and Wachter (1994). They also used mobility and poverty which are Census 2000 long form data items not available at this time.

At the A.C.E block cluster level, within post-strata, one can construct an indicator of total coverage, the coverage gap, as follows:

$$z = \text{(weighted P-sample non-matches)} - \text{(weighted E-sample erroneous enumerations)}$$

At the block cluster level, a correlation between z and each artificial population's estimated true net coverage error (see Appendix for details) can be made. Note that each artificial population uses two surrogate variables, one for gross undercount and one for gross overcount. Because of the possibly large amount of geocoding error at the block cluster level, these correlations will likely be small. Large correlations may merely mean that our artificial populations are related to geocoding error. Whatever the case, the correlations may be used to help rank the artificial populations in order of importance.

From this analysis, multiple sets of artificial populations are selected for calculation of the error of synthetic estimates.

Bias due to synthetic estimation

The bias of a synthetic estimate can be split into two components:

- the synthetic population bias due to carrying the post-stratum level net coverage adjustment down to the state and congressional district levels
- bias in the DSE including correlation bias

The first component is estimated from an artificial population; it is the synthetic estimate minus the population count estimated from the artificial population.

The second component is estimated using post-stratum biases, estimated as part of the Total Error Model. It is the post-stratum level biases in the DSE allocated to the state and congressional district levels.

The estimated bias for shares accounts for the same two components of error as for levels.

See the Appendix for detailed formulas.

Bias in census counts

The bias in the census count for an area is the census count minus the population count estimated from the artificial population.

Results

What are the results of the artificial population creation?

Based on the block cluster level correlation analysis, four artificial populations were created as described in Table 1. Among all the combinations of overcount and undercount surrogates considered, these were the four that had the highest correlations. Artificial population 4 had the highest correlation among potential artificial populations that excluded remainder surrogates (i.e. excludes surrogates formed by subtracting the number of persons with a characteristics such as substituted from the total number of persons). Typical correlations obtained ranged from slightly negative to around 0.26.

Table 1: Surrogate Variables used to Create Artificial Populations

	Correlations (weighted analysis)	Undercount Surrogate	Overcount Surrogate
Artificial Population 1	**0.26**	**# non-substituted persons in households**	**#persons for whom reported date of birth and reported age were consistent (allocation not required)**
Artificial Population 2	**0.27**	**# non-substituted persons in households**	**# non-substituted persons in households**
Artificial Population 3	**0.26**	**# persons with 2 or more items allocated**	**#persons for whom reported date of birth and reported age were consistent (allocation not required)**
Artificial Population 4	**0.25**	**# persons whose household did not mail back the questionnaire**	**# persons whose household did not mail back the questionnaire**

Household Persons only (Group Quarters Persons are Excluded)

Note that for Artificial Populations 2 and 4 the same surrogate variable is used for undercount and overcount. Thus if the post-strata has an overall undercount (overcount) all local areas will have an undercount (overcount) for that post-strata for these artificial populations. See the Appendix for details.

Regional examples of artificial population creation

Tables 2a, 3a, 4a, and 5a illustrate the creation of the four artificial population counts at the regional level. The actual artificial populations are created at the congressional district level and summed to the state and region levels. Thus, these illustrations are not exactly equal to what is obtained by summing over the congressional districts but they are very close.

For each table the total U.S. gross undercount is allocated to the regions in proportion to the undercount surrogate variable. The total U.S. gross overcount is allocated to the regions in proportion to the overcount surrogate variable. The artificial population count is then given by: census count + allocated gross undercount - allocated gross overcount.

Tables 2b, 3b, 4b, and 5b below show the census error and the synthetic error for each of the four artificial populations at the regional level. The census error is the census count minus the artificial population count. The synthetic error is the synthetic estimate minus the artificial population count. From A.C.E. the gross estimates of undercount and overcount are:

- U.S. Gross Undercount = 16,296 (in thousands)
- U.S. Gross Overcount = 13,034 (in thousands)

What are levels of the components of bias in synthetic estimates for states?

Tables 6, 7, 8, and 9 give the components of bias in the synthetic estimates at the State level for Artificial Populations 1, 2, 3, and 4 respectively. Columns (1) through (4) are for estimates of level. Column (1), SynB, is the synthetic population bias and column (2), DSEB, is the DSE level bias including correlation bias. Column (3) is the ratio of SynB to DSEB. Column (4) is the relative total bias in the state level synthetic estimate of level. Column (5) is the bias in the estimate of share. Column (6) is the relative bias in the synthetic estimate of population share.

The results for Artificial Populations 1 and 2 are similar. Artificial Population 4 is similar to these on average but has more variation. Artificial Population 3 is different than the others.

How does bias in synthetic estimates compare to bias in the census?

For a given state, let absolute census bias be defined as the absolute value of the census count (or share) minus the true count (or share) from the artificial population. Similarly let the absolute synthetic bias be defined as the absolute value of the synthetic estimate of count (or share) minus the true count (or share) from the artificial population. Next define the ratio, R, of the absolute census bias to the absolute synthetic bias, denoted

$$R = \frac{|census - true|}{|synthetic - true|} .$$

Since R does not indicate if the errors are large relative to the true value, define the statistic relative error, as follows:

$$RELERR = \frac{|census - true| - |synthetic - true|}{true} .$$

RELERR is approximately the same for counts and shares.

At the state level, how does the bias in the synthetic estimates compare with the bias in the census numbers?

Tables 10 and 11 show the state level percentiles of the ratio R for the artificial populations. At the tails of the distributions of the ratios for shares, the values are quite small (or large) because the census (or the synthetic estimate) is very close to the true value as measured by the artificial population. For each artificial population for both counts and shares, synthetic estimation improved the count for the majority of states (the median ratio is greater than 1 for all cases). The percentiles of RELERR are also shown. The absolute relative error median is less than 0.6 percent for each artificial population for both counts and shares.

At the congressional district level, how does the total bias in the synthetic estimates compare with the bias in the census numbers?

Tables 12 and 13 show the percentiles of the ratio R for the artificial populations for congressional districts. At the tails of the distributions of the ratios for shares and counts, the values are quite small (or large) because the census (or the synthetic estimate) is very close to the true value as measured by the artificial population. For each artificial population for both counts and shares, synthetic estimation improved the count for the majority of states (the median ratio is greater than 1 for all cases). The percentiles of RELERR are also shown. The absolute relative error median is less than 0.5 percent for each artificial population for both counts and shares.

What is the effect of synthetic error on the unweighted squared error loss function analysis?

The loss function analysis does not include an error component for the failure of the synthetic assumption. An expression for a bias correction to a squared error loss function difference, Loss(Census) - Loss(A.C.E.), is shown in the Appendix. This bias correction term can be added to loss function results to correct for the bias of excluding synthetic error in the loss function target estimates. The interpretation of the bias correction term is most relevant in terms of the sign of the squared error loss function difference. If the loss function difference is positive, indicating adjustment is favorable, only a negative bias correction can change this making adjustment unfavorable. Similarly, if the difference is negative, indicating adjustment is not favorable, this can be reversed only if the bias correction is positive. The amount of bias being added or subtracted must be larger than the absolute difference to reverse the outcome.

Tables 14 through 17 show the bias correction term for states and congressional districts for estimated counts and estimated shares. In each table results are shown for each of the four artificial populations. Column (1) is the census squared error loss minus the adjusted squared error loss. This has a bias due to excluding synthetic error. Column (2) is the synthetic bias correction term. Column (3) is the relative bias (column (2) / column (1)). Column (4) is the bias corrected loss function difference (column (1) + column (2)).

For state level count estimates (Table 14), three of the four artificial populations show a positive correction in the census loss minus the A.C.E. loss to correct this bias. Thus, the loss function analysis is conservative. In other words, for these three artificial populations, the loss function analysis is underestimating the true gains from adjustment. For the other artificial populations the bias correction is negative but is only 20.79 percent of the Census loss minus the A.C.E. loss. Thus, correcting for the bias would not change the loss function results.

For state level share estimates (Table 15), three of the four artificial populations show a negative bias correction. For two of these cases this negative bias correction is less than 6 percent. However, even for the negative 55.67 percent bias, the loss function results are not reversed. The loss function analysis is conservative for the remaining artificial population.

For congressional district count estimates (Table 16), three of the four artificial populations show a negative bias correction is necessary. However, in all three of these cases this negative bias correction is less than 8 percent of the difference in census and A.C.E. Thus, correcting for the bias would not reverse the loss function results. For the other artificial population, the loss function analysis is conservative.

For congressional district share estimates (Table 17), the bias correction is positive for two of the four artificial populations, and results of the loss function analysis go from being favorable to the census to being favorable to A.C.E. For the other two artificial populations, the bias correction is negative and large (75.86 percent and 31.79 percent). For these cases the loss function analysis result was favorable to the census and the correction would not change this although the result would be stronger.

What is the effect of synthetic error on the equal CD squared error loss function?

Table 18 provides the same results for the Equal CD Squared Error Loss Function.

For congressional district share estimates using the Equal CD Squared Error Loss Function (Table 18), the bias correction is positive for one of the four artificial populations, and the loss function analysis is conservative. For the other three artificial populations, the bias correction is negative but less in absolute value than the difference in the loss function analysis (column (1)). Thus, the outcome of the loss function analysis would not be reversed.

What is the effect of synthetic error on the weighted squared error loss function analysis?

Tables 19 through 22 are similar to Tables 14 through 17, respectively, but a weighted squared error loss function analysis is used. For both states and congressional districts, the weight is 1 / (census count).

For state level count estimates (Table 19), all of the artificial populations have a negative bias correction. However each of these bias corrections has an absolute value less than 6 percent of the loss function analysis difference (column (1)). Thus, correcting for the bias would not change the loss function results.

For state level share estimates (Table 20), all of the artificial populations have a negative bias correction. For the first three artificial populations the bias correction has an absolute value less than 8 percent of the loss function analysis difference (column (1)). Artificial Population 4 has an absolute relative bias of 57.69 percent which is not negligible but much less than the loss function difference. Thus, correcting for the bias would not change the loss function results for any of the artificial populations.

For congressional district count estimates (Table 21), three of the four artificial populations show a negative bias correction is necessary. However, in all three of these cases this negative bias correction is less than 8 percent of the difference in census and A.C.E. Thus, correcting for the bias would not reverse the loss function results. For the other artificial population, the loss function analysis is conservative.

For congressional district share estimates (Table 22), the bias correction is positive for two of the four artificial populations and the loss function analysis is conservative. For the other two artificial populations, the bias correction is negative (12.04 percent and 3.7 percent) but much smaller in absolute value than the loss function difference. Thus correcting for the bias would not reverse the loss function results.

What set of bias estimates from the total error model and loss function analysis were used for examining the effect of synthetic bias on loss function analysis?

All loss function results cited in this report use the model which includes correlation bias except for Non-Blacks ages 18-29 and uses the Gross DSE to distribute target estimates. Using Gross Undercount to distribute target estimates keeping the correlation bias assumption fixed would produce results of similar magnitude and sign. We did not run alternative correlation bias assumptions; we think these results are reasonable under these alternatives but we are not completely confident of this. See Navarro and Asiala (2001) for information on how results differ with different DSE bias assumptions.

References

Fay, R.E. and J. Thompson (1993). "The 1990 Post Enumeration Survey Statistical Lessons in Hindsight." *Proceedings of the 1993 Annual Research Conference*. U.S. Bureau of the Census, 71-91.

Freedman, D. and K. Wachter (1994). "Heterogeneity and Census Adjustment for the Intercensal Base." *Statistical Science*, 476-485.

Hengartner, N. and T.P. Speed (1993). "Assessing between-block heterogeneity within the post-strata of the 1990 Post Enumeration Survey." *Journal of the American Statistical Association*, 88, 1047-1057.

Kim, J.J., A. Zaslavsky, and R. Blodgett (1995). "Between-State Heterogeneity of Undercount Rates and Surrogate Variables in the 1990 U.S. Census." *Survey Methodology*, 21, 1, pp.55-62.

Navarro, A. and M. Asiala, (2001). "Accuracy and Coverage Evaluation: Comparing Accuracy." DSSD Census 2000 Procedures and Operations Memorandom, Series B-13*, February 28, 2001.

Table 2a : Illustration of Artificial Population 1 Creation at Regional Level

region	census count (%) (1)	undercount surrogate (%) (2)	overcount surrogate (%) (3)	allocated undercount (4)	allocated overcount (5)	artificial pop. estimate (1)+(4)-(5)
Northeast	51,927 (19.0%)	51,149 (18.9%)	50,403(19.0%)	3,085	2,474	52,538
Midwest	62,601 (22.9%)	62,010 (23.0%)	61,063 (23.0%)	3,741	2,997	63,344
South	97,400 (35.6%)	96,112 (35.6%)	94,600 (35.6%)	5,798	4,643	98,554
West	61,659 (22.5%)	60,875 (22.5%)	59,477 (22.4%)	3,672	2,919	62,412
Total	273,587 (100%)	270,147 (100%)	265,543 (100%)	16,296	13,034	276,849

(Numbers in Thousands)

Table 2b: Regional Census Error and Synthetic Error using Artificial Population 1

region	census count	synthetic estimate	artificial pop. estimate	census error	synthetic error
Northeast	51,927	52,450	52,538	-611	-88
Midwest	62,601	63,007	63,344	-743	-337
South	97,400	98,789	98,554	-1,154	235
West	61,659	62,602	62,412	-753	190
Total	273,587	276,849	276,849	-3,262	0

(Numbers in Thousands)

Table 3a : Illustration of Artificial Population 2 Creation at Regional Level

region	census count (%) (1)	undercount surrogate (%) (2)	overcount surrogate (%) (3)	allocated undercount (4)	allocated overcount (5)	artificial pop. estimate (1)+(4)-(5)
Northeast	51,927 (19.0%)	51,149 (18.9%)	51,149 (18.9%)	3,085	2,468	52,544
Midwest	62,601 (22.9%)	62,010 (23.0%)	62,010 (23.0%)	3,741	2,992	63,350
South	97,400 (35.6%)	96,112 (35.6%)	96,112 (35.6%)	5,798	4,637	98,561
West	61,659 (22.5%)	60,875 (22.5%)	60,875 (22.5%)	3,672	2,937	62,394
Total	273,587 (100%)	270,147 (100%)	270,147 (100%)	16,296	13,034	276,849

(Numbers in Thousands)

Table 3b: Regional Census Error and Synthetic Error using Artificial Population 2

region	census count	synthetic estimate	artificial pop. estimate	census error	synthetic error
Northeast	51,927	52,450	52,544	-618	-94
Midwest	62,601	63,007	63,350	-749	-343
South	97,400	98,789	98,561	-1,161	229
West	61,659	62,602	62,394	-735	208
Total	273,587	276,849	276,849	-3,262	0

(Numbers in Thousands)

Table 4a: Illustration of Artificial Population 3 Creation at Regional Level

region	census count (%) (1)	undercount surrogate (%) (2)	overcount surrogate (%) (3)	allocated undercount (4)	allocated overcount (5)	artificial pop. estimate (1)+(4)-(5)
Northeast	51,927 (19.0%)	2,323 (19.0%)	50,403 (19.0%)	3,099	2,474	52,552
Midwest	62,601 (22.9%)	2,256 (18.5%)	61,063 (23.0%)	3,010	2,997	62,613
South	97,400 (35.6%)	4,407 (36.1%)	94,600 (35.6%)	5,880	4,643	98,637
West	61,659 (22.5%)	3,228 (26.4%)	59,477 (22.4%)	4,306	2,919	63,046
Total	273,587(100%)	12,214 (100%)	265,543 (100%)	16,296	13,034	276,849

(Numbers in Thousands)

Table 4b: Regional Census Error and Synthetic Error using Artificial Population 3

region	census count	synthetic estimate	artificial pop. estimate	census error	synthetic error
Northeast	51,927	52,450	52,552	-625	-102
Midwest	62,601	63,007	62,613	-12	393
South	97,400	98,789	98,637	-1,237	152
West	61,659	62,602	63,046	-1,387	-444
Total	273,587	276,849	276,849	-3,262	0

(Numbers in Thousands)

Table 5a: Illustration of Artificial Population 4 Creation at Regional Level

region	census count (%) (1)	undercount surrogate (%) (2)	overcount surrogate (%) (3)	allocated undercount (4)	allocated overcount (5)	artificial pop. estimate (1)+(4)-(5)
Northeast	51,927 (19.0%)	7,732 (16.3%)	7,732 (16.3%)	2,654	2,123	52,458
Midwest	62,601 (22.9%)	9,924 (20.9%)	9,924 (20.9%)	3,407	2,725	63,283
South	97,400 (35.6%)	18,564 (39.1%)	18,564 (39.1%)	6,373	5,098	98,676
West	61,659 (22.5%	11,245 (23.7%)	11,245 (23.7%)	3,861	3,088	62,432
Total	273,587 (100%)	47,465 (100%)	47,465 (100%)	16,296	13,034	276,849

(Numbers in Thousands)

Table 5b: Regional Census Error and Synthetic Error using Artificial Population 4

region	census count	synthetic estimate	artificial estimate	census error	synthetic error
Northeast	51,927	52,450	52,458	-531	-8
Midwest	62,601	63,007	63,283	-682	-276
South	97,400	98,789	98,676	-1,276	114
West	61,659	62,602	62,432	-773	170
Total	273,587	276,849	276,849	-3,262	0

(Numbers in Thousands)

Table 6: State Level Synthetic Bias Using Artificial Population 1

State	SynB (1)	DSEB (2)	SynB/DSEB (3)	(SynB+DSEB)/N (4)	B-share (5)	rel. B-share (6)
Alabama	1395	40,534	0.0344	0.0096	0.000037	0.0023
Alaska	-337	5,757	-0.0586	0.0096	0.000005	0.0023
Arizona	413	34,663	0.0119	0.0069	-0.000006	-0.0003
Arkansas	-481	27,496	-0.0175	0.0103	0.000029	0.0030
California	-10067	189,122	-0.0532	0.0053	-0.000227	-0.0019
Colorado	-464	27,684	-0.0168	0.0064	-0.000012	-0.0008
Connecticut	-728	21,183	-0.0344	0.0061	-0.000013	-0.0011
Delaware	475	4,483	0.1060	0.0064	-0.000002	-0.0008
D.C.	148	3,616	0.0409	0.0069	-0.000001	-0.0003
Flordia	-1483	91,291	-0.0162	0.0057	-0.000087	-0.0015
Georgia	-414	67,634	-0.0061	0.0083	0.000032	0.0011
Hawaii	145	8,937	0.0162	0.0076	0.000001	0.0003
Idaho	-90	10,220	-0.0088	0.0079	0.000003	0.0007
Illinois	2563	92,868	0.0276	0.0078	0.000027	0.0006
Indiana	2568	44,218	0.0581	0.0079	0.000014	0.0006
Iowa	-416	24,639	-0.0169	0.0085	0.000013	0.0013
Kansas	-382	22,185	-0.0172	0.0083	0.000010	0.0011
Kentucky	-377	33,811	-0.0112	0.0084	0.000017	0.0012
Louisiana	124	39,118	0.0032	0.0089	0.000027	0.0017
Maine	-420	11,978	-0.0351	0.0092	0.000009	0.0020
Maryland	2895	28,527	0.1015	0.0060	-0.000023	-0.0012
Massachusetts	-1404	39,612	-0.0354	0.0062	-0.000023	-0.0010
Michigan	-2228	61,939	-0.0360	0.0061	-0.000038	-0.0011
Minnesota	-193	34,760	-0.0056	0.0072	-0.000000	-0.0000
Mississippi	-385	30,977	-0.0124	0.0110	0.000038	0.0037
Missouri	-843	43,858	-0.0192	0.0079	0.000013	0.0007
Montana	-484	10,844	-0.0446	0.0116	0.000014	0.0044
Nebraska	-298	13,349	-0.0223	0.0078	0.000004	0.0006
Nevada	715	15,252	0.0469	0.0080	0.000006	0.0008
New Hampshire	315	8,934	0.0353	0.0076	0.000002	0.0004
New Jersey	-928	54,644	-0.0170	0.0065	-0.000023	-0.0008
New Mexico	920	19,416	0.0474	0.0112	0.000026	0.0039
New York	10838	158,134	0.0685	0.0091	0.000125	0.0019
North Carolina	-1227	71,077	-0.0173	0.0088	0.000046	0.0016
North Dakota	-6	6,392	-0.0009	0.0103	0.000007	0.0030
Ohio	-3561	67,350	-0.0529	0.0057	-0.000059	-0.0015
Oklahoma	-1140	30,578	-0.0373	0.0087	0.000018	0.0015
Oregon	-699	17,415	-0.0401	0.0049	-0.000028	-0.0023
Pennsylvania	-73	83,376	-0.0009	0.0070	-0.000010	-0.0002
Rhode Island	429	7,348	0.0583	0.0076	0.000002	0.0004
South Carolina	295	29,295	0.0101	0.0075	0.000005	0.0003
South Dakota	-1	7,801	-0.0001	0.0107	0.000009	0.0034
Tennessee	191	40,272	0.0047	0.0072	-0.000000	-0.0000
Texas	6888	161,275	0.0427	0.0081	0.000069	0.0009
Utah	-1096	13,785	-0.0795	0.0057	-0.000012	-0.0015
Vermont	15	5,698	0.0026	0.0096	0.000005	0.0023
Virginia	518	47,847	0.0108	0.0070	-0.000006	-0.0002
Washington	-1711	26,599	-0.0643	0.0043	-0.000062	-0.0029
West Virginia	-739	20,328	-0.0364	0.0110	0.000024	0.0037
Wisconsin	676	34,693	0.0195	0.0067	-0.000009	-0.0005
Wyoming	148	4,989	0.0296	0.0105	0.000006	0.0033
Average				0.0079		0.0007
Standard Deviation				0.0017		0

Table 7: State Level Synthetic Bias Using Artificial Population 2

State	SynB	DSEB	SynB/DSEB	(SynB+DSEB)/N	B-share	rel. B-share
	(1)	(2)	(3)	(4)	(5)	(6)
Alabama	200	40,534	0.0049	0.0093	0.000033	0.0021
Alaska	-142	5,757	-0.0246	0.0099	0.000005	0.0027
Arizona	456	34,663	0.0131	0.0069	-0.000006	-0.0003
Arkansas	-190	27,496	-0.0069	0.0104	0.000030	0.0031
California	-1972	189,122	-0.0104	0.0056	-0.000197	-0.0016
Colorado	177	27,684	0.0064	0.0066	-0.000010	-0.0007
Connecticut	-77	21,183	-0.0036	0.0063	-0.000010	-0.0009
Delaware	147	4,483	0.0329	0.0060	-0.000003	-0.0012
D.C.	46	3,616	0.0128	0.0067	-0.000001	-0.0005
Flordia	319	91,291	0.0035	0.0058	-0.000080	-0.0014
Georgia	120	67,634	0.0018	0.0084	0.000034	0.0012
Hawaii	-20	8,937	-0.0022	0.0074	0.000001	0.0002
Idaho	-36	10,220	-0.0035	0.0079	0.000003	0.0007
Illinois	724	92,868	0.0078	0.0077	0.000020	0.0005
Indiana	458	44,218	0.0104	0.0075	0.000006	0.0003
Iowa	-82	24,639	-0.0033	0.0087	0.000015	0.0014
Kansas	-95	22,185	-0.0043	0.0084	0.000011	0.0012
Kentucky	-260	33,811	-0.0077	0.0084	0.000017	0.0012
Louisiana	-115	39,118	-0.0029	0.0089	0.000026	0.0017
Maine	-61	11,978	-0.0051	0.0095	0.000010	0.0022
Maryland	647	28,527	0.0227	0.0056	-0.000031	-0.0016
Massachusetts	-75	39,612	-0.0019	0.0064	-0.000018	-0.0008
Michigan	-405	61,939	-0.0065	0.0063	-0.000032	-0.0009
Minnesota	-70	34,760	-0.0020	0.0072	0.000000	0.0000
Mississippi	-190	30,977	-0.0061	0.0111	0.000038	0.0038
Missouri	-220	43,858	-0.0050	0.0080	0.000015	0.0008
Montana	-89	10,844	-0.0082	0.0121	0.000016	0.0048
Nebraska	-91	13,349	-0.0068	0.0079	0.000004	0.0007
Nevada	194	15,252	0.0127	0.0077	0.000004	0.0005
New Hampshire	103	8,934	0.0115	0.0074	0.000001	0.0002
New Jersey	-7	54,644	-0.0001	0.0066	-0.000019	-0.0006
New Mexico	140	19,416	0.0072	0.0108	0.000023	0.0035
New York	2144	158,134	0.0136	0.0086	0.000094	0.0014
North Carolina	-392	71,077	-0.0055	0.0089	0.000049	0.0017
North Dakota	-31	6,392	-0.0049	0.0102	0.000007	0.0030
Ohio	-681	67,350	-0.0101	0.0060	-0.000049	-0.0012
Oklahoma	-354	30,578	-0.0116	0.0089	0.000021	0.0017
Oregon	-92	17,415	-0.0053	0.0051	-0.000026	-0.0021
Pennsylvania	-29	83,376	-0.0003	0.0070	-0.000010	-0.0002
Rhode Island	146	7,348	0.0198	0.0074	0.000001	0.0001
South Carolina	37	29,295	0.0013	0.0075	0.000004	0.0003
South Dakota	-29	7,801	-0.0037	0.0106	0.000009	0.0034
Tennessee	-111	40,272	-0.0028	0.0072	-0.000001	-0.0001
Texas	523	161,275	0.0032	0.0078	0.000046	0.0006
Utah	-258	13,785	-0.0188	0.0061	-0.000009	-0.0011
Vermont	42	5,698	0.0074	0.0096	0.000005	0.0024
Virginia	-122	47,847	-0.0026	0.0069	-0.000008	-0.0003
Washington	-195	26,599	-0.0073	0.0045	-0.000056	-0.0027
West Virginia	-161	20,328	-0.0079	0.0113	0.000026	0.0040
Wisconsin	3	34,693	0.0001	0.0066	-0.000011	-0.0006
Wyoming	26	4,989	0.0052	0.0103	0.000005	0.0031
Average				0.0079		0.0007
Standard Deviation				0.0017		0

Table 8: State Level Synthetic Bias Using Artificial Population 3

State	SynB	DSEB	SynB/DSEB	(SynB+DSEB)/N	B-share	rel. B-share
	(1)	(2)	(3)	(4)	(5)	(6)
Alabama	-20,429	40,534	-0.5040	0.0046	-0.000042	-0.0026
Alaska	9,433	5,757	1.6386	0.0273	0.000040	0.0199
Arizona	-26,995	34,663	-0.7788	0.0015	-0.000105	-0.0057
Arkansas	7,223	27,496	0.2627	0.0132	0.000057	0.0060
California	-65,950	189,122	-0.3487	0.0037	-0.000428	-0.0035
Colorado	-1,859	27,684	-0.0672	0.0061	-0.000017	-0.0011
Connecticut	[illegible],556	21,183	0.1679	0.0074	0.000003	0.0002
Delaware	-6,139	4,483	-1.3693	-0.0021	-0.000026	-0.0093
D.C.	-10,761	3,616	-2.9763	-0.0128	-0.000040	-0.0199
Flordia	-46,968	91,291	-0.5145	0.0028	-0.000251	-0.0044
Georgia	-78,861	67,634	-1.1660	-0.0014	-0.000251	-0.0085
Hawaii	7,676	8,937	0.8590	0.0139	0.000029	0.0066
Idaho	6,100	10,220	0.5969	0.0128	0.000025	0.0055
Illinois	-47,307	92,868	-0.5094	0.0037	-0.000154	-0.0035
Indiana	-24,875	44,218	-0.5626	0.0032	-0.000085	-0.0039
Iowa	6,621	24,639	0.2687	0.0110	0.000039	0.0038
Kansas	-713	22,185	-0.0322	0.0082	0.000009	0.0010
Kentucky	29,132	33,811	0.8616	0.0159	0.000124	0.0087
Louisiana	18,483	39,118	0.4725	0.0132	0.000093	0.0059
Maine	-9,615	11,978	-0.8027	0.0019	-0.000024	-0.0053
Maryland	-6,434	28,527	-0.2255	0.0042	-0.000056	-0.0030
Massachusetts	2,786	39,612	0.0703	0.0069	-0.000008	-0.0003
Michigan	3,881	61,939	0.0627	0.0067	-0.000016	-0.0005
Minnesota	-163	34,760	-0.0047	0.0072	-0.000000	-0.0000
Mississippi	-10,476	30,977	-0.3382	0.0073	0.000001	0.0001
Missouri	18,607	43,858	0.4243	0.0115	0.000083	0.0042
Montana	6,479	10,844	0.5975	0.0196	0.000039	0.0123
Nebraska	11,078	13,349	0.8299	0.0147	0.000045	0.0075
Nevada	-6,371	15,252	-0.4177	0.0044	-0.000020	-0.0028
New Hampshire	-6,873	8,934	-0.7693	0.0017	-0.000024	-0.0055
New Jersey	2,588	54,644	0.0474	0.0069	-0.000010	-0.0003
New Mexico	7,436	19,416	0.3830	0.0148	0.000049	0.0076
New York	-1,846	158,134	-0.0117	0.0084	0.000079	0.0012
North Carolina	21,200	71,077	0.2983	0.0117	0.000127	0.0045
North Dakota	3,359	6,392	0.5255	0.0158	0.000019	0.0085
Ohio	48,031	67,350	0.7132	0.0104	0.000127	0.0032
Oklahoma	24,028	30,578	0.7858	0.0162	0.000109	0.0090
Oregon	17,524	17,415	1.0062	0.0104	0.000038	0.0031
Pennsylvania	-23,313	83,376	-0.2796	0.0050	-0.000094	-0.0022
Rhode Island	-1,068	7,348	-0.1453	0.0062	-0.000004	-0.0010
South Carolina	15,320	29,295	0.5229	0.0114	0.000059	0.0042
South Dakota	3,725	7,801	0.4775	0.0159	0.000023	0.0086
Tennessee	17,913	40,272	0.4448	0.0104	0.000064	0.0032
Texas	37,131	161,275	0.2302	0.0096	0.000178	0.0024
Utah	13,427	13,785	0.9740	0.0123	0.000040	0.0051
Vermont	-4,460	5,698	-0.7828	0.0021	-0.000011	-0.0051
Virginia	61,915	47,847	1.2940	0.0160	0.000216	0.0087
Washington	-17,647	26,599	-0.6634	0.0015	-0.000119	-0.0056
West Virginia	8,534	20,328	0.4198	0.0162	0.000057	0.0090
Wisconsin	4,387	34,693	0.1264	0.0075	0.000005	0.0002
Wyoming	1,550	4,989	0.3107	0.0135	0.000011	0.0062
Average				0.0086		0.0014
Standard Deviation				0.0065		0

Table 9: State Level Synthetic Bias Using Artificial Population 4

State	SynB	DSEB	SynB/DSEB	(SynB+DSEB)/N	B-share	rel. B-share
	(1)	(2)	(3)	(4)	(5)	(6)
Alabama	-3,236	40,534	-0.0798	0.0085	0.000020	0.0013
Alaska	2,771	5,757	0.4813	0.0151	0.000016	0.0079
Arizona	-5,669	34,663	-0.1635	0.0057	-0.000028	-0.0015
Arkansas	-1,200	27,496	-0.0436	0.0100	0.000026	0.0027
California	20,832	189,122	0.1102	0.0063	-0.000115	-0.0009
Colorado	-355	27,684	-0.0128	0.0064	-0.000012	-0.0008
Connecticut	1,339	21,183	0.0632	0.0068	-0.000005	-0.0004
Delaware	-1,874	4,483	-0.4179	0.0034	-0.000011	-0.0038
D.C.	2,999	3,616	0.8295	0.0121	0.000010	0.0049
Flordia	-21,247	91,291	-0.2327	0.0044	-0.000158	-0.0028
Georgia	7,515	67,634	0.1111	0.0093	0.000061	0.0021
Hawaii	2,334	8,937	0.2612	0.0094	0.000009	0.0022
Idaho	-409	10,220	-0.0400	0.0076	0.000002	0.0004
Illinois	-46,240	92,868	-0.4979	0.0038	-0.000150	-0.0034
Indiana	-16,104	44,218	-0.3642	0.0047	-0.000054	-0.0025
Iowa	1,142	24,639	0.0464	0.0091	0.000019	0.0019
Kansas	-3,612	22,185	-0.1628	0.0071	-0.000001	-0.0001
Kentucky	360	33,811	0.0107	0.0086	0.000020	0.0014
Louisiana	-6,757	39,118	-0.1727	0.0074	0.000002	0.0001
Maine	-2,478	11,978	-0.2069	0.0075	0.000001	0.0003
Maryland	-3,017	28,527	-0.1058	0.0049	-0.000044	-0.0023
Massachusetts	7,211	39,612	0.1820	0.0076	0.000008	0.0004
Michigan	-3,273	61,939	-0.0528	0.0060	-0.000042	-0.0012
Minnesota	2,303	34,760	0.0663	0.0077	0.000009	0.0005
Mississippi	1,668	30,977	0.0538	0.0117	0.000045	0.0045
Missouri	-4,968	43,858	-0.1133	0.0071	-0.000002	-0.0001
Montana	1,394	10,844	0.1285	0.0138	0.000021	0.0065
Nebraska	-76	13,349	-0.0057	0.0079	0.000004	0.0007
Nevada	2,003	15,252	0.1313	0.0086	0.000010	0.0014
New Hampshire	-2,703	8,934	-0.3025	0.0051	-0.000009	-0.0021
New Jersey	10,763	54,644	0.1970	0.0079	0.000020	0.0007
New Mexico	-227	19,416	-0.0117	0.0106	0.000022	0.0033
New York	72,415	158,134	0.4579	0.0124	0.000347	0.0052
North Carolina	-10,190	71,077	-0.1434	0.0077	0.000014	0.0005
North Dakota	3,074	6,392	0.4809	0.0153	0.000018	0.0080
Ohio	-10,850	67,350	-0.1611	0.0051	-0.000085	-0.0021
Oklahoma	-1,463	30,578	-0.0479	0.0086	0.000017	0.0014
Oregon	-8,687	17,415	-0.4988	0.0026	-0.000057	-0.0046
Pennsylvania	8,682	83,376	0.1041	0.0077	0.000021	0.0005
Rhode Island	-394	7,348	-0.0536	0.0068	-0.000001	-0.0004
South Carolina	-5,430	29,295	-0.1853	0.0061	-0.000016	-0.0011
South Dakota	1,867	7,801	0.2394	0.0133	0.000016	0.0060
Tennessee	-9,880	40,272	-0.2453	0.0054	-0.000036	-0.0018
Texas	17,830	161,275	0.1106	0.0087	0.000108	0.0015
Utah	269	13,785	0.0195	0.0063	-0.000007	-0.0009
Vermont	-2,539	5,698	-0.4457	0.0053	-0.000004	-0.0019
Virginia	7,662	47,847	0.1601	0.0080	0.000020	0.0008
Washington	-5,346	26,599	-0.2010	0.0036	-0.000075	-0.0036
West Virginia	7,295	20,328	0.3589	0.0155	0.000053	0.0083
Wisconsin	-4,038	34,693	-0.1164	0.0058	-0.000026	-0.0014
Wyoming	-1,465	4,989	-0.2937	0.0072	-0.000000	-0.0000
Average				0.0079		0.0007
Standard Deviation				0.0030		0

Table 10: State Level Percentiles for Statistics Comparing Absolute Census Bias to Absolute Synthetic Bias - Artificial Populations 1 and 2

Percentile	Count Artificial Population 1	Count Artificial Population 2	Share Artificial Population 1	Share Artificial Population 2	RELERR Artificial Population 1	RELERR Artificial Population 2
5	0.527	0.530	0.165	0.182	-0.004	-0.004
10	0.745	0.718	0.398	0.427	-0.002	-0.002
25	1.120	1.130	0.971	0.889	0.001	0.001
50	1.500	1.520	1.990	2.380	0.005	0.005
75	2.060	2.070	6.740	7.680	0.008	0.008
90	2.610	2.490	10.760	14.500	0.009	0.010
95	2.890	2.910	28.820	23.970	0.014	0.014

Table 11: State Level Percentiles for Statistics Comparing Absolute Census Bias to Absolute Synthetic Bias- Artificial Populations 3 and 4

Percentile	Count Artificial Population 3	Count Artificial Population 4	Share Artificial Population 3	Share Artificial Population 4	RELERR Artificial Population 3	RELERR Artificial Population 4
5	0.067	0.310	0.078	0.312	-0.015	-0.008
10	0.228	0.562	0.190	0.536	-0.011	-0.005
25	0.439	1.130	0.540	0.895	-0.005	0.001
50	1.040	1.530	1.200	1.990	0.000	0.006
75	3.610	2.190	2.100	5.600	0.011	0.009
90	10.710	3.100	11.110	14.470	0.018	0.010
95	11.180	4.130	23.540	26.390	0.021	0.013

Table 12: CD Level Percentiles for Statistics Comparing Absolute Census Bias to Absolute Synthetic Bias - Artificial Populations 1 and 2

Percentile	Count Artificial Population 1	Count Artificial Population 2	Share Artificial Population 1	Share Artificial Population 2	RELERR Artificial Population 1	RELERR Artificial Population 2
5	0.426	0.410	0.074	0.115	-0.005	-0.005
10	0.662	0.650	0.189	0.232	-0.002	-0.003
25	1.060	1.080	0.723	0.766	0.000	0.001
50	1.560	1.570	2.130	2.060	0.004	0.004
75	2.320	2.300	4.780	4.660	0.008	0.008
90	3.530	3.360	12.810	11.480	0.012	0.012
95	4.160	3.870	26.290	25.500	0.015	0.014

Table 13: CD Level Percentiles for Statistics Comparing Absolute Census Bias to Absolute Synthetic Bias- Artificial Populations 3 and 4

Percentile	Count Artificial Population 3	Count Artificial Population 4	Share Artificial Population 3	Share Artificial Population 4	RELERR Artificial Population 3	RELERR Artificial Population 4
5	0.073	0.278	0.147	0.095	-0.015	-0.008
10	0.141	0.524	0.328	0.230	-0.012	-0.005
25	0.399	0.928	0.665	0.597	-0.007	0.000
50	1.400	1.620	1.440	1.600	0.003	0.004
75	3.990	2.720	2.960	3.730	0.013	0.009
90	11.530	5.660	6.400	7.680	0.021	0.014
95	22.750	11.050	12.650	18.680	0.025	0.018

Table 14: Loss Function Synthetic Bias Correction for State Counts

	Squared Error Loss			
Artificial Population	Census Loss minus A.C.E. Loss (1)	Synthetic Bias Correction (2)	Relative Bias (3)	Corrected Loss (4)
1	3.01E+11	9.16E+08	0.30%	3.02E+11
2	3.01E+11	5.58E+08	0.19%	3.01E+11
3	3.01E+11	9.09E+10	30.24%	3.92E+11
4	3.01E+11	-6.25E+10	-20.79%	2.38E+11

Table 15: Loss Function Synthetic Bias Correction for State Shares

	Squared Error Loss			
Artificial Population	Census Loss minus A.C.E. Loss (1)	Synthetic Bias Correction (2)	Relative Bias (3)	Corrected Loss (4)
1	3.03E-07	-1.68E-08	-5.57%	2.86E-07
2	3.03E-07	-2.32E-11	-0.008%	3.03E-07
3	3.03E-07	4.08E-08	13.48%	3.43E-07
4	3.03E-07	-1.68E-07	-55.67%	1.34E-07

Table 16: Loss Function Synthetic Bias Correction for Congressional District Counts

	Squared Error Loss			
Artificial Population	Census Loss minus A.C.E. Loss (1)	Synthetic Bias Correction (2)	Relative Bias (3)	Corrected Loss (4)
1	1.33E+10	-3.82E+08	-2.87%	1.29E+10
2	1.33E+10	-7.51E+07	-0.57%	1.32E+10
3	1.33E+10	3.97E+09	29.93%	1.73E+10
4	1.33E+10	-1.00E+09	-7.54%	1.23E+10

Table 17: Loss Function Synthetic Bias Correction for Congressional District Shares

Squared Error Loss				
Artificial Population	Census Loss minus A.C.E. Loss (1)	Synthetic Bias Correction (2)	Relative Bias (3)	Corrected Loss (4)
1	-4.27E-06	-3.24E-06	75.86%	-7.51E-06
2	-4.27E-06	-1.36E-06	31.79%	-5.63E-06
3	-4.27E-06	6.29E-05	-1470.99%	5.86E-05
4	-4.27E-06	2.30E-05	-538.88%	1.88E-05

Table 18: Equal CD Loss Function Synthetic Bias Correction for Congressional District Shares

Weighted Squared Error Loss (Weight = square of state census count)				
Artificial Population	Census Loss minus A.C.E. Loss (1)	Synthetic Bias Correction (2)	Relative Bias (3)	Corrected Loss (4)
1	1.61E+09	-2.14E+08	-13.27%	1.40E+09
2	1.61E+09	-4.60E+07	-2.85%	1.57E+09
3	1.61E+09	2.86E+09	177.17%	4.47E+09
4	1.61E+09	-4.84E+08	-29.98%	1.13E+09

Table 19: Weighted Loss Function Synthetic Bias Correction for State Counts

Weighted Squared Error Loss (Weight = 1/ census count)				
Artificial Population	Census Loss minus A.C.E. Loss (1)	Synthetic Bias Correction (2)	Relative Bias (3)	Corrected Loss (4)
1	1.79E+04	-126.95	-0.71%	1.77E+04
2	1.79E+04	-6.07	-0.03%	1.79E+04
3	1.79E+04	-1.60	-0.01%	1.79E+04
4	1.79E+04	-990.00	-5.54%	1.69E+04

Table 20: Weighted Loss Function Synthetic Bias Correction for State Shares

Weighted Squared Error Loss (Weight = 1/ census count)				
Artificial Population	Census Loss minus A.C.E. Loss (1)	Synthetic Bias Correction (2)	Relative Bias (3)	Corrected Loss (4)
1	5.92E-06	-4.38E-07	-7.40%	5.48E-06
2	5.92E-06	-2.09E-08	-0.35%	5.90E-06
3	5.92E-06	-5.53E-09	-0.09%	5.91E-06
4	5.92E-06	-3.41E-06	-57.69%	2.50E-06

Table 21: Weighted Loss Function Synthetic Bias Correction for Congressional District Shares

Weighted Squared Error Loss (Weight = 1/ census count)				
Artificial Population	Census Loss minus A.C.E. Loss (1)	Synthetic Bias Correction (2)	Relative Bias (3)	Corrected Loss (4)
1	2.07E+04	-4.99E+02	-2.41%	2.02E+04
2	2.07E+04	-8.69E+01	-0.42%	2.06E+04
3	2.07E+04	5.64E+03	27.22%	2.64E+04
4	2.07E+04	-1.61E+03	-7.79%	1.91E+04

Table 22: Weighted Loss Function Synthetic Bias Correction for Congressional District Shares

Weighted Squared Error Loss (Weight = 1/ census count)				
Artificial Population	Census Loss minus A.C.E. Loss (1)	Synthetic Bias Correction (2)	Relative Bias (3)	Corrected Loss (4)
1	2.09E-04	-2.51E-05	-12.04%	1.84E-04
2	2.09E-04	-7.73E-06	-3.70%	2.01E-04
3	2.09E-04	4.99E-04	238.79%	7.07E-04
4	2.09E-04	3.83E-05	18.36%	2.47E-04

APPENDIX

Forming artificial populations

Let X denote a surrogate for gross undercount and Y denote a surrogate for gross overcount.

DSE_j = the Dual System Estimate for Post-stratum j

E_j = the weighted E sample total in post-stratum j

CE_j = the weighted E sample number of correct enumerations in post-stratum j

EE_j = the weighted E sample number of erroneous enumerations in post-stratum j

$Cen_{.j}$ = the census count in post-stratum j

Note that for any variable V, $V_{.j}$ is the sum of V_{ij} over areas i.

Define the estimated gross undercount as follows:

$$GUNDER_j = DSE_j - Cen_{.j}(\frac{CE_j}{E_j})$$

Define the estimated gross overcount as follows:

$$GOVER_j = Cen_{.j}(\frac{EE_j}{E_j})$$

N_{ij} is the artificial population value and Cen_{ij} is the census count for area i, post-stratum j.

$$N_{ij} = Cen_{ij} + X_{ij}\frac{GUNDER_j}{X_{.j}} - Y_{ij}\frac{GOVER_j}{Y_{.j}}$$

$$N_{.j} = Cen_{.j} + GUNDER_j - GOVER_j = Cen_{.j} + DSE_j - Cen_{.j} = DSE_j$$

The artificial populations were selected by computing the, within post-strata, correlation between

z=(Weighted P-sample Non-matches)- (Weighted E-sample erroneous enumerations).

and $N_{ij} - Cen_{ij}$, at the A.C.E. block cluster level.

Decomposition of the Error in a Synthetic Estimate into Two Additive Components.

Notation

N_i = the true population for area i

cen_{ij} = census count for area i, post-stratum j

$cen_{.j}$ = census count in post-stratum j

$CF_{.j} = \frac{N_{.j}}{cen_{.j}}$ = true coverage correction factor for post-stratum j

$\hat{CF}_{.j} = \frac{DSE_{.j}}{cen_{.j}}$ = estimated coverage factor for post-stratum j

$\hat{N}_{i.} = \sum_j \hat{CF}_{.j}\, cen_{ij}$ = the A.C.E. synthetic estimate for area i

$\tilde{N}_{i.} = \sum_j CF_{.j}\, cen_{ij}$ = the known population synthetic estimate for area i

Then $\hat{N}_i - N_i = (\tilde{N}_{i.} - N_i) + (\hat{N}_i - \tilde{N}_i)$

Define:

$B_i = E(\hat{N}_i - N_{i.})$, the bias in the synthetic estimate

$SynB_i = \tilde{N}_i - N_i$, the error due to carrying down the true post-stratum coverage correction factors to area i. Since the true coverage correction factors are used, bias in the DSE at the post-stratum level is excluded from this error.

$DSEB_i = E(\hat{N}_i - \tilde{N}_i)$, the error due to using the estimated coverage correction factors instead of the true coverage correction factors for each post-stratum. This error is due to bias in the DSE including correlation bias.

Specifying Bias due to Synthetic Estimation

The first component of the synthetic bias is estimated using artificial populations, the second component is estimated using post-stratum biases, estimated as part of the Total Error Model and Loss Function work.. The estimate of bias for area i takes the following form:

$$\hat{B}_i = Sy\hat{n}B_i + DS\hat{E}B_i = (\tilde{N}_i - N_i) + \sum_j \frac{Cen_{ij}}{Cen_{.j}} \hat{D}_j .$$

Here, the first part is estimated from an artificial population; it is the known artificial population synthetic count (equivalent to the production synthetic estimate because the artificial populations are adjusted so that the total over areas for a post-stratum equals the DSE) minus the actual population count from the artificial population.

The second part contains the post-stratum bias, $\hat{D}_j$, (estimated elsewhere) which is an estimate of: ($E(DSE_j)$-the true population of post-stratum j). The true population of post-stratum j is estimated using results from the Total Error Model Analysis. In this second term, we weight the post-stratum bias by the proportion of post-stratum census counts in area i .

2. <u>The bias of the synthetic estimator of share</u>.

The bias for the synthetic estimator of a population share for area i takes the following form:

$$\hat{B}_{share,\, i} = \frac{N_i + Sy\hat{n}B_i + DS\hat{E}B_i}{\sum_i (N_i + Sy\hat{n}B_i + DS\hat{E}B_i)} - \frac{N_i}{\sum_i N_i}$$

Correction for Synthetic Bias in Loss Function Analysis

Notation:

D_g = the census squared error loss minus the A.C.E. squared error loss using synthetic target estimates.

D_t = the census squared error loss minus the A.C.E. squared error loss using "true" target estimates

The loss function analysis output is in terms of expected losses using the synthetic target estimates, i.e., $\Delta_g = E(D_g)$. However, we would like to know $\Delta_t = E(D_t)$. Therefore, we develop

an expression for a bias correction term, B, to be added to Δ_g to correct loss function results for synthetic bias so that

$$\Delta_t = \Delta_g + B.$$

Define:

w_i = the squared error loss function weight for area i (set equal to 1 for an unweighted squared error loss)

Cen_i = the census count for area i

N_i = the "true" target estimate for area i

$\tilde{N}_i$ = the synthetic target estimate for area i

$\hat{N}_i$ = the A.C.E. synthetic estimate for area i (including DSE post-stratum biases)

b_i = bias in the post-stratum level DSE including correlation bias allocated to area i

By definition,

$$a_i = E(\hat{N}_i) = \tilde{N}_i + b_i$$

Using this notation:

$$D_g = \sum_i [w_i(Cen_i - \tilde{N}_i)^2 - w_i(\hat{N}_i - \tilde{N}_i)^2], \text{ and}$$

$$D_t = \sum_i [w_i(Cen_i - N_i)^2 - w_i(\hat{N}_i - N_i)^2]$$

$$= D_g + 2\sum_i w_i(\tilde{N}_i - N_i)(Cen_i - \hat{N}_i)$$

The resulting expected difference is:

$$\Delta_t = \Delta_g + 2\sum_i w_i(\tilde{N}_i - N_i)(Cen_i - a_i)$$

$$= \Delta_g + 2\sum_i w_i(\tilde{N}_i - N_i)(Cen_i - \tilde{N}_i - b_i),$$

so B = bias correction term = $2\sum_i w_i(\tilde{N}_i - N_i)(Cen_i - \tilde{N}_i - b_i)$.

Estimates for this bias term are made by using artificial population values for the terms N_i and $\tilde{N}_i$ and by estimating b_i with $\sum_j \frac{Cen_{ij}}{Cen_j}\hat{D}_j$. An analogous approach is used for shares.